This is an author version before the final review. You can find the final version here: https://doi.org/10.1016/j.comnet.2018.02.012

Please cite as: Cetto, A.; Klier, M.; Richter, A.; Zolitschka, J.F. (2018): "Thanks for sharing"—Identifying users' roles based on knowledge contribution in Enterprise Social Networks. Computer Networks, Volume 135, pp. 275-288

"Thanks for Sharing" – Identifying Users' Roles Based on Knowledge Contribution in Enterprise Social Networks

Alexandra Cetto, University of Regensburg, Universitätsstr. 31, 93055 Regensburg, Germany, Alexandra.Cetto@ur.de

Prof. Dr. Mathias Klier, University of Ulm, Helmholtzstr. 22, 89081 Ulm, Germany, Mathias.Klier@uni-ulm.de (primary contact)

Prof. Dr. Alexander Richter, IT-University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen S, Denmark, aric@itu.dk

Felix Zolitschka, University of Ulm, Helmholtzstr. 22, 89081 Ulm, Germany, Felix.Zolitschka@uni-ulm.de

Abstract

While ever more companies use Enterprise Social Networks for knowledge management, there is still a lack of understanding of users' knowledge exchanging behavior. In this context, it is important to be able to identify and characterize users who contribute and communicate their knowledge in the network and help others to get their work done. In this paper, we propose a new methodological approach consisting of three steps, namely "message classification", "identification of users' roles" as well as "characterization of users' roles". We apply the approach to a dataset from a multinational consulting company, which allows us to identify three user roles based on their knowledge contribution in messages: givers, takers, and matchers. Going beyond this categorization, our data shows that whereas the majority of messages aims to share knowledge, matchers, that means people that give and take, are a central element of the network. In conclusion, the development and application of a new methodological approach allows us to contribute to a more refined understanding of users' knowledge exchanging behavior in Enterprise Social Networks which can ultimately help companies to take measures to improve their knowledge management.

Keywords

Enterprise Social Networks, User roles, Knowledge contribution, Knowledge sharing, Knowledge seeking

This is an author version before the final review. You can find the final version here: https://doi.org/10.1016/j.comnet.2018.02.012

Please cite as: Cetto, A.; Klier, M.; Richter, A.; Zolitschka, J.F. (2018): "Thanks for sharing"—Identifying users' roles based on knowledge contribution in Enterprise Social Networks. Computer Networks, Volume 135, pp. 275-288

1 Introduction

It is forecast that the global market for Enterprise Social Networks (ESN) will grow by a 19% average year-on-year which means the annual revenue will hit $3.5 billion by 2019 [1]. ESN are online platforms used in a business context, which facilitate light-weight communication via short messages and provide functions to find, connect, and interact with colleagues [2]. With their ability to offer large-scale benefits in enterprise communication, collaboration, knowledge sharing, and thus organizational knowledge management [3–5], they are gaining rapid adoption [6]. The intention behind is to improve communication practices as well as business agility through an enhanced employee engagement [7,8]. As a matter of fact, ESN can create competitive advantage favoring effective and efficient business [9,10].

First studies have shown that these social technologies can support knowledge practices like information seeking, knowledge sharing or expert finding [11–13]. In this context, users' rationales behind online knowledge sharing [e.g., 14,15] and knowledge seeking [e.g., 16–18] as well as the relationships between both aspects [14,19] have already been investigated. Moreover, prior studies have identified users' roles based on the users' knowledge contribution behavior in ESN. These studies focus mainly on the users' structural positions and characteristics in networks [e.g., 20–22].

However, in this context there is still a missing understanding of how different users and their roles impact ESN usage and how the underlying network structures influence information dissemination and contribution behavior [18]. More specifically, to the best of our knowledge identifying users' roles based on the contents exchanged in an ESN has not yet been subject of academic discussion. This perspective bears huge potential as about 80% of an organization's information is contained within text documents [23] and there are calls to deepen the understanding of these potentials [e.g., 24,25]. Therefore, our objective in this paper is to regard the knowledge practices of ESN users from a content perspective and identify users' roles based on their knowledge contribution in contents, in particular in messages. As a consequence, we address the following research questions:

1. How can users' roles in ESN be identified based on users' knowledge contribution in messages?
2. How can the users be characterized depending on their roles and the messages that they exchanged with other users?

In order to answer these questions, we suggest a new methodological approach consisting of three steps namely "message classification", "identification of users' roles" and "characterization of users' roles". We apply the approach to a large volume of ESN communication data from a multinational consulting company using the ESN Yammer, to come to a better understanding of the characteristics of the messages written and the users connected on the platform. A text analysis approach allows us to classify messages as "knowledge sharing" and "knowledge seeking". On this basis, we identify users depending on their knowledge sharing and seeking behavior in messages, i.e. their knowledge contribution in the ESN. In detail, we identify givers (users who outstandingly share knowledge), takers (users who outstandingly seek knowledge), and matchers (users who share and seek knowledge to a relatively balanced extent and therefore are in-between the two extremes of givers and takers). In addition, we analyze the structural characteristics of the users via Social Network Analysis and further activities in the network as well as the characteristics of the messages written by users of each user role.

Our results indicate that the majority of messages aim at sharing knowledge as compared to seeking knowledge. Moreover, most users contribute knowledge for others but also expect information in return which identifies them as matchers. Those users are also amongst the best connected users which gives them a central position in the ESN.

Our contribution to theory and practice is first of all the new methodological approach to analyze users' knowledge exchanging behavior and its application to ESN data from a multinational consulting company: (1) We distinguish ESN users' messages based on their knowledge sharing and seeking content via text analysis. (2) We identify users as givers, takers, and matchers based on their knowledge contribution, which depends on their knowledge sharing and seeking messages. (3) We investigate the characteristics of the users' roles such as typical structural positions in the network and content patterns or lengths of their messages. Moreover, the application of our approach reveals results that contribute to a more refined understanding of ESN usage and can ultimately help companies to improve their knowledge management.

The remainder of this paper is structured as follows: We first review the existing literature on knowledge sharing and seeking as well as users' roles based on knowledge contribution in ESN and identify the research gap. We then describe the context of our case study as well as the analyzed data and provide insights into the used method. Afterwards, we present our results, followed by a discussion. We conclude with our contribution and an outlook on future research.

2 Background and Related Work

In recent years, organizations discovered the potential of ESN to facilitate corporation-wide knowledge exchange without being subject to departmental or geographic boundaries [2] and contribute to more open and participative communication practices [26–28]. By now, ESN are often considered a crucial means for companies to stay competitive [3]. Research about ESN already covered the adoption of ESN in organizations [29], the development of relationships between employees [18,30], the potential benefits of ESN in the corporate realm, including expert finding, problem solving, work coordination, and opinion sharing [12,31,32] as well as ESN's influence on career paths [33] and the relationship between ESN and formal hierarchies [34,35]. Yet, there is an increasing call to better understand the behavior of users in ESN [36,37], especially in reference to information diffusion [38] and knowledge exchange [39,40].

2.1 Knowledge Sharing and Seeking in Enterprise Social Networks

Sharing with others and demanding in return is a natural behavior pattern of mankind. When interacting with each other, individuals have to decide between the two extremes whether to claim as much value as possible or contribute value without expecting anything in return [41]. Reciprocity has been considered as one of the most important factors that determine individuals' knowledge contribution or sharing behaviors in online communities [42,43]. The consideration of knowledge as public good allows that knowledge exchange is driven by care for the community rather than by self-interest [44]. Hence, a person who has gained something from someone else tends to give something back in return in order to sustain ongoing supportive exchanges [45]. Against this background, reciprocity becomes a dominant determinant of knowledge sharing behavior [46,47].

On a related note and following the increasing demand to better understand the users' knowledge contribution in ESN, research started to investigate knowledge sharing and seeking of ESN users. There is a significant body of research showing that enterprise social software, such as ESN, is used for knowledge exchanging rather than for socializing [e.g., 48,49]. Hence, employees engage with the aim of searching and finding new corporate knowledge, which shows that the value for the employee is rather based on information-gathering as opposed to social purposes [50]. Jackson et al. [48] and Thom-Santelli et al. [51] find that users of corporate tagging and blogging systems aim at providing information and being thought as leaders rather than seeking information for themselves or connecting with colleagues for social purposes.

In this line of research, in particular the individuals' rationales behind online knowledge sharing [e.g., 14,15] and seeking [e.g., 16–18] as well as the relation between online sharing and seeking knowledge [14,19] are investigated. Concerning the knowledge exchanging behavior in ESN, Wasko and Faraj [17] investigate the reasons of some users contributing more than others. They find that users do so if they notice an enhancement of their professional reputation, enjoy helping others, are structurally embedded in the network, and/or if their experiences are worth sharing with others. They further identify the obligation of reciprocity, i.e. giving back to the community in return for help, as drivers of knowledge sharing behavior [52]. In this context, Nowak and Sigmund [53] and Mathews and Green [54] state that reciprocity derives from the desire to repay the help or knowledge received from the community before. The relevance of reciprocity for knowledge sharing has also been confirmed by other studies [e.g., 42, 43,55]. Kankanhalli et al. [16] find that knowledge self-efficacy and enjoyment in helping others significantly impact knowledge contribution to electronic repositories whereas the loss of knowledge power and image do not appear to have any impact. Zhang and Wang [18] state that a person's position in the network influences the decisions about his or her total contribution as well as the allocation of his or her efforts on the platform. Schroer and Hertel [15] refer to contributions in an encyclopedia and find the predictors of contributors' engagement and satisfaction to be determined by perceived benefits, identification with the community, and task characteristics. Besides, their engagement depends on their tolerance for opportunity costs and the experienced characteristics of their tasks, which again is partially mediated by intrinsic motivation.

Further studies regard knowledge sharing and seeking in ESN from the content perspective and analyze the contents shared and exchanged within ESN as sources of knowledge. Riemer and Richter [56] explore communication patterns in ESN text messages by applying manual text analysis to 648 posts and find that the texts can be classified in different genres, such as "Ask questions", "Share information", or "Discuss and clarify". They conclude that communication in their case is targeted towards providing awareness information for others and coordinating task and team matters. Cleveland [57] states that social networks in the corporate context enable users to re-post texts of other users in their own network which makes sharing knowledge with new audiences possible. They therefore allow for capturing and transferring project knowledge in organizations and facilitate the conversations between users for the purpose of sharing lessons [58]. Zhang et al. [59] investigate an ESN at a Fortune 500 company in a five-month study and find that the platform is mainly used to share information through messages with specially formed groups that particularly engage in long conversations, which in turn facilitates knowledge sharing among the employees. They show that users can more easily build connections, find answers to specific questions, and that the informal communication is improved. In an approach to classify text documents, Ebner et al. [60] conduct a study by tracking students' messages in an ESN which was used for communication, collaboration, and documentation during a course. Of a total of 11,214 posts which were manually assigned to pre-defined categories, 60% could be identified as reply posts, indicating a clear communication process between users. These results indicate potential for informal learning and project-oriented communication on the platform. In keeping with this, Zhao et al. [61] examine the virtual network communication of a large IT company and find that 91% of the 886 posts were work-relevant, more precisely 44% were associated with tasks statuses, 19% with information and idea sharing, 18% with other work-related statuses, and 6% with questions.

Indeed, while contents exchanged in ESN have already been analyzed in prior research, to date the content perspective has not been subject to research concerning users' knowledge sharing and seeking behavior in ESN in particular. We assume that it is essential to not only investigate the rationales behind and relationships between knowledge sharing and seeking in ESN but to likewise consider the content perspective in particular in messages when analyzing users' knowledge contribution. Thus, the users' knowledge contributions to an ESN are determined based on their knowledge sharing and seeking messages using text classification algorithms. Indeed, in this context automated approaches are needed due to the rising popularity of ESN and thus the rising amount of written messages available.

2.2 Users' Roles Based on Knowledge Contribution in Enterprise Social Networks

Social scientists state that people differ tremendously in their preferences for reciprocity – their desired mix of giving and taking. Against this backdrop, Grant [41] classifies people as givers (i.e. people who give more than they get), takers (i.e. people who get more than they give) and matchers (i.e. people who try to trade evenly). Also in terms of knowledge contribution in ESN not all users can be considered as equal [e.g., 20–22]. They differ, for instance, regarding the contents they produce with respect to frequency, volume, and quality [62]. To analyze the users and their roles more in depth, Grant's framework can serve as a starting point as it helps to differentiate between people with preferences for sharing knowledge or seeking knowledge in ESN.

People's roles in terms of knowledge contribution have been analyzed in the context of knowledge work. Knowledge work is rooted in the transformation of the society into a post-industrial state where work shifted from being manual to non-manual. The feature differentiating knowledge work from other conventional work is that its basic task is thinking [63]. Hence, knowledge workers' primary purpose involves the creation, distribution, and application of knowledge [64]. Among the roles are central connectors, boundary spanners or peripheral specialists [e.g., 65]. Reinhardt et al. [66] identify ten knowledge worker roles depending on their knowledge sharing and seeking actions and propose among others controller, helper, learner, linker, networker or sharer as roles.

Former research in the context of users' roles in ESN already addresses the users' structural positions in the network and finds that only a few individuals receive a majority of the attention in ESN [20]. Furthermore, there is often a small number of very active users as contrasted with a large number of rather passive users, so called lurkers [21,22,67]. Nonnecke and Preece [68] find that the share can range to as much as 99% of the users and point out that there are different reasons for lurking in online social networks, with usability problems or reluctance being examples [69]. Schneider et al. [70] draw the connection between epistemic curiosity as personality trait and emotional-motivational state to lurkers' contribution in online communities and reveal that the psychology of curiosity generally holds great promise for research on online communities in information systems.

Understanding why users share and seek knowledge is especially important with respect to ESN, as users largely differ in terms of connectivity (e.g. number of friends), communication activity (e.g. number of messages) as well as frequency, volume, and quality of the user-generated content [62]. In this regard, Gloor et al. [71] analyze users' contributions in networks based on their communication patterns. They examine the users' contribution index, i.e. the extent to which their communication is balanced between sending and receiving messages, and are able to identify leadership roles. Trier and Richter [72] identify two different and interrelated actor roles as an explanation for uneven levels of user contributions to ESN. They call them discourse drivers and information retrievers as two mutually interdependent actors, which together shape the dynamics of the online interaction. On a related note, Beck et al. [25] identify knowledge contributors and knowledge seekers as two interconnected user roles in terms of knowledge exchange by analyzing their characteristics as well as their dyadic relationship from an activity-centered language/action point of view. Holtzblatt et al. [73] differentiate between active contributors, moderate contributors, and readers as well as active and occasional users while analyzing log data. Moreover, Berger et al. [74] focus on users' structural aspects and coin the term of value adding users. They find that the users who add value to the organization by sharing their knowledge in the ESN are amongst the best connected users and thus enable a more effective and rapid exchange of information between different working groups. Additionally, Cetto et al. [24] investigate knowledge sharing and seeking of ESN users in a knowledge base and identify givers, takers, and matchers based on their number of write and read accesses.

With our study, we build upon these approaches. While the majority of studies concerns solely users' structural characteristics, we want to focus in particular on the contents exchanged and identify users' roles based on knowledge contribution considering the content perspective in addition to the analysis of users' structural aspects. In detail, we aim at identifying givers, takers, and matchers in ESN based on their knowledge sharing and seeking in messages. While most research on user roles in ESN focusses on solely two contrary user roles, only few researchers also define user roles being in-between these extremes. Holtzblatt et al. [73] for example shape the term of moderate contributors as user role between active contributors and readers. Reflecting the research on knowledge work, indeed roles being mixes of contrary roles were proposed. Among them are linker, networker, and helper. Linkers are described as "people who associate and mash up information from different sources to generate new information" while networkers are "people who create personal or project related connections with people involved in the same kind of work, to share information and support each other" [66, p. 11]. Helpers can be defined as "people who transfer information once they passed a problem" [66, p. 11]. Nevertheless, these roles do not entirely explain the meaning of matchers and thus are not applicable for the purpose of our study. For this reason as well as for consistency reasons we decided to stay with the term matcher to ensure an adequate analysis.

3 Research Gap

Multiple authors emphasize the benefits of social media like ESN for knowledge practices [11–13,75]. At the same time there is still a missing understanding of how different users and their roles impact ESN usage and how the underlying network structures influence information dissemination and contribution behavior [18]. Previous studies recognized that users differ in their contribution in ESN and identified users' roles according to their activities and structural characteristics in the network. Cetto et al. [24] provide a first step towards the identification of givers, takers, and matchers. However, they only focused on the users' structural characteristics and the mere number of write and read accesses in the ESN, but did not take into consideration the content perspective. As 80% of an organization's information is contained within texts [23], such as messages, it bears huge potential to also consider the contents within ESN as sources of knowledge. Furthermore, content analysis have already been conducted in comparable network contexts for the purpose of analyzing contents as sources of knowledge [e.g., 56]. In addition, related research already called for further analyses of the contents when investigating knowledge contribution in networks [e.g., 24]. Therefore, we assume that investigating users' roles based on knowledge contribution in contents harbors enormous potential for research about knowledge contribution in ESN.

Hence, we build upon the approach of givers, takers, and matchers based on knowledge contribution in ESN and enrich this research stream by adopting the content perspective for the identification of the users' roles. Against this background, to the best of our knowledge we are the first to identify givers, takers, and matchers based on their knowledge sharing and seeking in messages using text analysis.

4 Methodological Considerations

4.1 Case Context

The selected case organization is a large multinational consulting company, which employs more than 180,000 people in 40 countries worldwide. In September 2008, a small group of consultants at the Dutch division of the company started to use Yammer.com, a web-based platform launched in the same month.

Yammer is a cloud service and as of today is used by about 500,000 companies worldwide as well as of 85% of the Fortune 500[1]. The functionalities of Yammer are based on the "follower principle" where users choose whom they follow and can see by whom they are followed which can be seen on each user's profile. Another feature of Yammer is the opportunity to create groups with regard to certain topics which can be joined by users of the whole network and in which users can send messages to the group members

[1] https://products.office.com/de-de/yammer/yammer-overview

accordingly. Further platform features include profile information, options to send direct messages, and to like and bookmark posts.

The group of consultants was interested in creating an ESN environment to support their knowledge sharing by connecting employees with each other, creating more transparency, and making information easier to find. As such use practices were not yet established in the organization, the group of facilitators wanted to explore if and how the desired working practices could be established on the platform.

4.2 Dataset

The dataset arises from the first two years of Yammer usage in the company and was provided in MS Excel format for 10,432 unique users of the platform. 7,304 of these users followed at least one other user of the platform. To ensure confidentiality, all personal information (e.g. user names, email addresses) had been removed prior to handing over the data. The data contain 110,910 messages posted inside the ESN during the time period which were written by 9,806 users. Each message consists of metadata such as message ID, user ID, timestamp, and the content of the message. In Yammer, a message can either be sent to one person or a group as private message or as a public message which can be seen by the whole network accordingly. Excluding messages with no content, each message consists of a minimum of one and a maximum of 509 terms. 5,242 messages (4.73%) were sent to one recipient (direct communication), 35,273 messages (31.80%) were sent in groups, and 16,719 messages (15.07%) were private. Furthermore, the data comprises 14,946 likes in reply to messages that were sent by 984 users of the platform. In addition, the dataset includes 599 bookmarks which were stored by users for later retrieval and information about 282 groups which the users can become a member of.

4.3 Data Analysis

Our study aims at investigating the users' knowledge contribution in ESN to identify them as givers, takers, and matchers based on their knowledge sharing and seeking messages (cf. Figure 1). By this means, it intends to provide further insights on users' roles based on knowledge contribution in ESN. Those users are regarded as matchers who share and seek knowledge to a relatively balanced extent. Users who outstandingly share knowledge (as compared to matchers) are regarded as givers whereas users who outstandingly seek knowledge are regarded as takers.

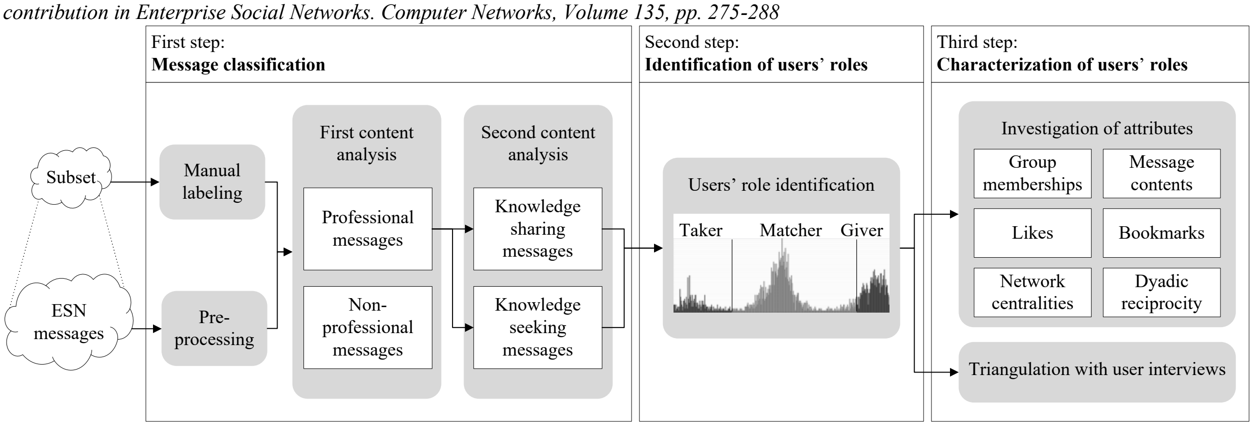

Figure 1 Research Approach

As a vast amount of company knowledge is contained within employees' written texts [23], we base our study on the users' knowledge contribution via ESN messages. Against this backdrop, we analyze messages in the ESN in a first step. Therefore, we classify messages as professional versus non-professional as merely the exchange of professional knowledge is relevant for the company's stock of knowledge. The professional messages are subsequently classified as knowledge sharing versus knowledge seeking messages. This serves us as a basis for identifying the users' roles based on their knowledge contribution in a second step. We identify each user as giver, taker, or matcher based on his or her previously identified knowledge sharing and knowledge seeking messages. To get deeper insights, we furthermore investigate in a third step the characteristics of the users' roles – such as typical structural positions in the network and content patterns – and finally triangulate the results with user interviews. Figure 1 provides an overview of our approach. It is composed of three steps which are discussed more in detail in the subsequent paragraphs.

4.3.1 First Step: Message Classification

In order to substantiate the definitions of givers, takers, and matchers for our setting, we first conduct text analysis to the messages written in the ESN. The aim is to find first, professional messages as a basis for further analyzing these professional messages in terms of their knowledge contributing content. For our study, only the professional messages are of interest as source of company relevant knowledge for further content analysis. As a common proceeding to identify relevant content in mass text-based messages, text analysis, consisting of the substeps *data preparation*, *data preprocessing*, *classification*, and *classifier evaluation,* is widely used as it has been proven to deliver reliable results [e.g., 76–78].

In text analysis, sample labeling is a critical step in order to train a classifier [e.g., 79,80]. Therefore, during the *data preparation* substep of our content analysis, we firstly construct a profile for each class. We define "professional" messages as containing information about the work in the company (e.g., technologies, directions, responsibilities, staffing) and/or about the network (Yammer) itself (e.g., functionalities). "Non-professional" message contents are regarded as non-informing or not work-related. Following this, we further subdivide the professional messages into the classes "knowledge sharing" and "knowledge seeking". In line with this, a message is regarded as "knowledge sharing" if it contains helpful information for other users (e.g., advices, helpful links, email addresses, references to documents or responsible persons), or if it offers help to other users. A "knowledge seeking" message in turn contains signs that the user receives information, demands for information, or demands help from other users. Hence, a team of two researchers manually code a randomly selected subset (training and test data) of 5% of the 110,910 messages to the corresponding class depending on the prevalence of the clear operational definitions above [81]. Therefore, each selected message is first coded to one of the two categories "professional" and "non-professional". Afterwards, if and only if the message is labelled "professional" it is further coded in "knowledge sharing" or "knowledge seeking". Regarding the coding procedure, to ensure reliable results, the researchers first define coding rules and label a first amount of messages together. After coding further 100 messages separately, Krippendorff alpha [82] as standard reliability measure for coding data [83] was used to estimate the inter-rater reliability and to ensure a consistent coding approach and reliable results. The two researchers reached a relatively high inter-coder reliability of α=0.8802. After consolidating the mismatches and refining the coding rules, each researcher codes by himself in order to reach a maximal subset of coded messages. We use 80% of this labelled data (training data) to train a classifier, utilize the remaining labelled data (test data) for classifier evaluation, and subsequently apply the classifier to the whole dataset.

In the following substep, *preprocessing* is used to clean the data and reduce the amount of terms to get the minimum of relevant terms to improve speed and accuracy of classification algorithms [84]. Preprocessing is composed of *feature extraction*, *feature representation*, and *feature selection*. *Feature extraction* is used to extract relevant features from the original text documents in a clear format [85]. Therefore, we remove all messages not relevant for our purpose, which are automatically generated messages (e.g. welcome notes and daily reports) identified by their standard structure and non-English messages identified by the Apache TIKA library[2]. Moreover, we conduct term manipulations to the texts in order to reduce the count of terms in such that we remove hyphen, markups from html, punctuations (except for question marks which are replaced by "questionmark" as they are assumed to be relevant for the identification of knowledge seeking messages), diacritic marks and numbers. Moreover, we write all terms to lower case. We also replace terms consisting of hyperlinks, emails, tags, user names or groups by categorical identifiers (e.g. "ishyperlink", "istag") as they have the same semantic meanings for our analysis. Additionally, we remove stop words (such as "and" and "the") using the built-in list of English stop words in KNIME[3] and reduce the terms to

---

[2] https://tika.apache.org/1.14/detection.html

[3] https://www.knime.org/files/nodedetails/_labs_textprocessing_preprocessing_Stop_word_Filter.html

their word stem using the common Porter Stemmer [86]. The purpose of *feature selection* is to eliminate irrelevant and redundant information from the target texts using a score mostly based on the frequency of terms [87]. We decide to eliminate features with only a single occurrence as they are not relevant for the classification. To conduct the previously described substeps we apply the commonly used bag-of-word *feature representation* for the preprocessing steps [88]. The result of the *feature selection* is represented in a vector space model in which each dimension represents a separate term as a single word with each term occurring at least once in a certain minimum number of documents [89].

In the *classification* substep, we apply text classification algorithms to the document-term matrix to assign a document to the corresponding class. In the last substep *classifier evaluation*, we aim at finding the best classification results. To do so, we test the most common classification algorithms, such as decision tree classification [90], support vector machine (SVM) [91], k-nearest neighbor (KNN) [92], naïve-bayesian [93] and artificial neural network (ANN) [94]. We train each algorithm with the labeled training data and evaluate the classifiers based on the labeled test data. For the evaluation, we compute recall, precision, accuracy, and F1 score which are widely used to assess the results of text analysis [95] and other machine learning approaches [96]. We utilize a 10-fold cross-validation [97] and regard the performance measures accuracy and F1-score as they include the measures recall and precision. We choose SVM as algorithm as it delivers good results and performs best for the classification in "professional" and "non-professional". Moreover, it also delivers good results for the classification in "knowledge sharing" and "knowledge seeking" (cf. Table 1).

	First content analysis				***Second content analysis***			
	Recall	Precision	Accuracy	F1	Recall	Precision	Accuracy	F1
ANN	81.4%	**89.4%**	80.6%	85.2%	95.6%	85.4%	87.0%	90.2%
Decision Tree	85.5%	85.7%	80.2%	85.6%	89.0%	**89.3%**	86.4%	89.1%
KNN	67.9%	85.3%	69.9%	75.6%	53.1%	91.8%	67.8%	67.3%
Naïve Bayes	**95.7%**	80.1%	79.7%	87.2%	**99.1%**	70.2%	73.2%	82.2%
SVM	90.0%	88.5%	**85.1%**	**89.2%**	92.1%	89.1%	**88.0%**	**90.6%**

Table 1 Classification Results

4.3.2 Second Step: Identification of Users' Roles

To identify givers, takers, and matchers based on knowledge contribution in ESN we not only binarily differentiate the messages m_i^u of a user u into messages with preliminary knowledge sharing content $c_{\uparrow}(m_i^u)$ and preliminary knowledge seeking content $c_{\downarrow}(m_i^u)$ but rather we use the probabilistic outputs

This is an author version before the final review. You can find the final version here: https://doi.org/10.1016/j.comnet.2018.02.012

Please cite as: Cetto, A.; Klier, M.; Richter, A.; Zolitschka, J.F. (2018): "Thanks for sharing"—Identifying users' roles based on knowledge contribution in Enterprise Social Networks. Computer Networks, Volume 135, pp. 275-288

$P\left(c_{\uparrow}(m_i^u)\right), P\left(c_{\downarrow}(m_i^u)\right)$, delivered by the SVM classifier [98]. $P\left(c_{\uparrow}(m_i^u)\right)$ represents the probability that the message is knowledge sharing while $P\left(c_{\downarrow}(m_i^u)\right)$ depicts the probability that the message is knowledge seeking. In order to establish the link between the average probabilities of users' messages being knowledge sharing and seeking, we take the differences between the averages (based on each user's amount of messages n^u) of $P\left(c_{\uparrow}(m_i^u)\right)$ and $P\left(c_{\downarrow}(m_i^u)\right)$. This difference represents each user's knowledge contribution to the ESN $\left(d_{c_{\uparrow\downarrow}}(u)\right)$. The results of the following formula are values in the interval $[-1; 1]$.

$$d_{c_{\uparrow\downarrow}}(u) = \overbrace{\left(\frac{1}{n^u}\sum_{i=1}^{n^u} P\left(c_{\uparrow}(m_i^u)\right)\right)}^{probability\ of\ messages\ being\ knowledge\ sharing} - \underbrace{\left(\frac{1}{n^u}\sum_{i=1}^{n^u} P\left(c_{\downarrow}(m_i^u)\right)\right)}_{probability\ of\ messages\ being\ knowledge\ seeking} = \frac{1}{n^u}\left(\sum_{i=1}^{n^u}\left(P\left(c_{\uparrow}(m_i^u)\right) - P\left(c_{\downarrow}(m_i^u)\right)\right)\right)$$

Based on the definition of $d_{c_{\uparrow\downarrow}}(u)$, users with $d_{c_{\uparrow\downarrow}}(u) = 1$ and $d_{c_{\uparrow\downarrow}}(u)$ close to 1 are regarded a givers, while users with $d_{c_{\uparrow\downarrow}}(u) = -1$ and $d_{c_{\uparrow\downarrow}}(u)$ close to -1 are regarded as takers. Matchers are located between givers and takers with a relatively balanced knowledge contribution. We define matchers as users whose knowledge contribution $d_{c_{\uparrow\downarrow}}(u)$ differs less than one standard deviation $\left(\sigma_{d_{c_{\uparrow\downarrow}}(u)}\right)$ from the average knowledge contribution $\left(\bar{x}_{d_{c_{\uparrow\downarrow}}(u)}\right)$ of all users. This results in a corresponding interval of $\left(\bar{x}_{d_{c_{\uparrow\downarrow}}(u)} - \sigma_{d_{c_{\uparrow\downarrow}}(u)}, \bar{x}_{d_{c_{\uparrow\downarrow}}(u)} + \sigma_{d_{c_{\uparrow\downarrow}}(u)}\right)$ for the knowledge contribution $d_{c_{\uparrow\downarrow}}(u)$ of a matcher. All users with a knowledge contribution $d_{c_{\uparrow\downarrow}}(u)$ outside this interval are classified as givers or takers respectively (cf. Figure 2).

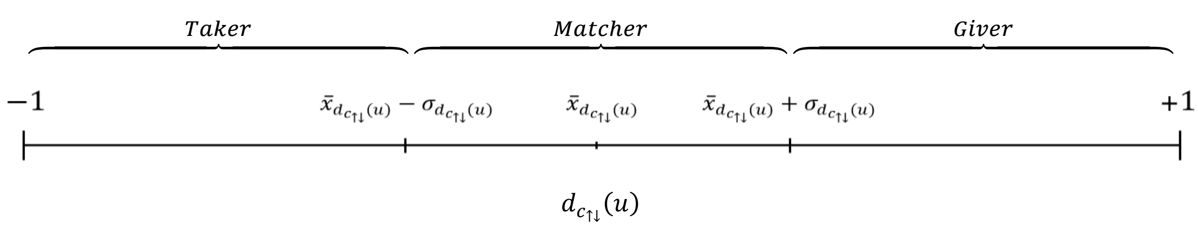

Figure 2 Identification of Users' Roles depending on $\boldsymbol{d_{c_{\uparrow\downarrow}}(u)}$

4.3.3 Third Step: Characterization of Users' Roles

To investigate the structural characteristics of givers, takers, and matchers and to analyze their social networking behavior in ESN, we apply Social Network Analysis [99]. In the context of ESN, Social Network Analysis was prior used to analyze for instance users' social networking behavior in ESN [35,100], or the characteristics of key users in ESN [74]. According to Freeman [101], Social Network Analysis "involves theorizing, model building, and empirical research focused on uncovering the patterning of links among actors" by, for instance, quantifying the centrality of nodes within a network via centrality measures. The most common centrality measures are degree centrality, closeness centrality, and betweenness centrality [102] as well as eigenvector centrality [103]. An ESN can be represented as a graph with a set of nodes (users) and a set of edges (ties) linking pairs of nodes. The edges can be directed or undirected and represent either social links like follower relations (social graph) or communication activities like messages amongst the users (activity graph) [104–106]. We base our research on both graphs in order to get profound insights into the structural characteristics of givers, takers, and matchers in ESN.

We analyze two different types of relationships: (1) social relationships (based on directed follower relations) and (2) communication (based on direct messages). In the case of *directed follower relations*, the social graph contains 9,237 nodes (users involved in at least one follower relationship) and 137,550 directed edges created by users who follow. Concerning *communication*, nodes represent senders and recipients of messages, while edges are created by sending a message. The corresponding activity graph contains 1,327 nodes (users being senders and/or receivers of at least one direct message in the observation period) and 2,660 directed edges. For our analysis, we use the igraph package for *R*[4] to calculate degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality for each node of the social as well as of the activity graph. In addition, we investigate a further network measure related to the users' reciprocal behavior. This measure refers to the behavior of responding to an action in the network with another action [107,108]. In ESN, reciprocal behavior shows up as retweeting, commenting, liking or answering one another's post or message to ensure ongoing mutual support. This reciprocity is regarded as dyadic level of analysis in terms of directed reciprocity [109]. We measure the users' dyadic level of knowledge contribution in ESN based on the users' sharing messages (to individuals as well as within groups), i.e. the percentage of answering with a sharing message to a previously received sharing message.

To gain further insights into the characteristics of givers, takers, and matchers, we additionally analyze user statistics. In particular, we investigate the content patterns and lengths of their messages exchanged and the participation of givers, takers, and matchers with respect to group memberships, usage of tags, bookmarks, and likes as well as messages received and sent (overall, private, public).

Finally, according to Behrendt et al. [110] combining results from different sources can improve the validity of the analysis. Thus, to draw a richer picture of the case context and complement our quantitative results, we decided to triangulate them with interviews with 14 users of the ESN which had been carried out earlier in order to get a better understanding of the ways users have appropriated the platform [111]. Yet, quotes

---

[4] http://cran.r-project.org/web/packages/igraph/index.html

from the interviews also proved helpful to illustrate how the ESN users reflected upon their behavior as well as on the behavior of others.

5 Results

This section is dedicated to the results. First, we focus on the results of the message classification. The second part concentrates on the results of the identification of users' roles and the third part reveals the characteristics of the users' roles in terms of structural positions in the network and content patterns as well as the triangulation with user interviews.

5.1 Results of the First Step: Message Classification

The results of the first step "message classification" reveal that professional messages are prevalent with a share of 72.4% in the ESN and exceed non-professional messages (27.6%) in many ways: they exhibit in total a vaster amount of messages, have a higher term variety as well as more terms per message as opposed to non-professional messages (cf. Table 2). Further, the majority of professional messages is classified as knowledge sharing with a share of 64.9% as opposed to knowledge seeking (33.1%). As to that, the knowledge sharing messages show higher amounts regarding term variety as well as terms per message indicating that knowledge sharing messages in average are longer and seem to contain more information than knowledge seeking messages (cf. Table 2).

	First content analysis		***Second content analysis***	
	Professional	Non-professional	Knowledge sharing	Knowledge seeking
No. of messages	57,056 (72.4%)	21,798 (27.6%)	38,194 (64.9%)	18,862 (33.1%)
Term variety	46,696	20,554	37,956	19,704
Terms/message	12.3	6.9	13.3	10.4

Table 2 Message Distribution and Attributes

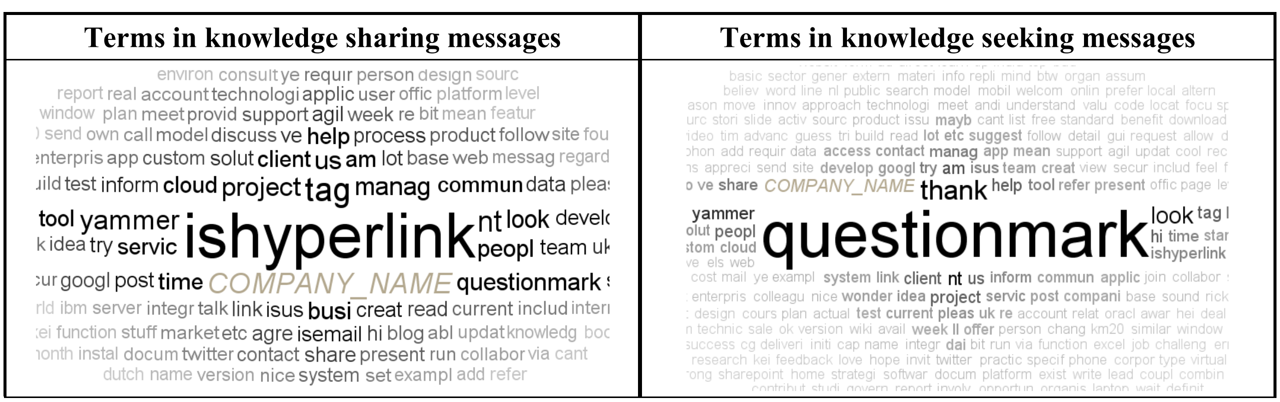

Figure 3 Word Clouds and Most Frequent Terms

Concerning the contents of the messages (cf. Figure 3), knowledge sharing messages include most of all the term "ishyperlink" (10,656 occurrences: e.g., "here you can read how you can do it ishyperlink") followed by "tag" (3,938 occurrences: e.g., "expecting quick market uptake of open group it specialist certification forthcoming year ishyperlink [istag] [istag]") and "n[o]t" (3,241 occurrences: e.g., "it is not a question of security but privacy"). Concerning knowledge seeking messages, the prevalent term is "questionmark" (14,836 occurrences: e.g., "any news on your potential project questionmark"), before "thank" (3,836 occurrences: e.g., "a really good ebook thanks for sharing"), and "looking[for]" (2,561 occurrences: e.g., "looking for reverences where we have done website rationalization"). The most frequent terms in knowledge sharing and seeking messages can be seen in the word clouds of Figure 3. At this point, it should be noticed that the same term can appear in knowledge sharing as well as knowledge seeking messages (i.e., "does this help someone questionmark" or "can someone help me questionmark").

5.2 Results of the Second Step: Identification of Users' Roles

As givers, takers, and matchers are identified based on their knowledge contribution in messages $d_{c_{\uparrow\downarrow}}(u)$, in the step "identification of users' roles" only users with knowledge sharing or knowledge seeking messages are regarded, which results in 2,734 users with at least one knowledge sharing or seeking message. The result of the identification of users' roles is depicted in Figure 4 which shows the distribution of $d_{c_{\uparrow\downarrow}}(u)$ among all users as well as the separation of givers, takers, and matchers according to the thresholds.

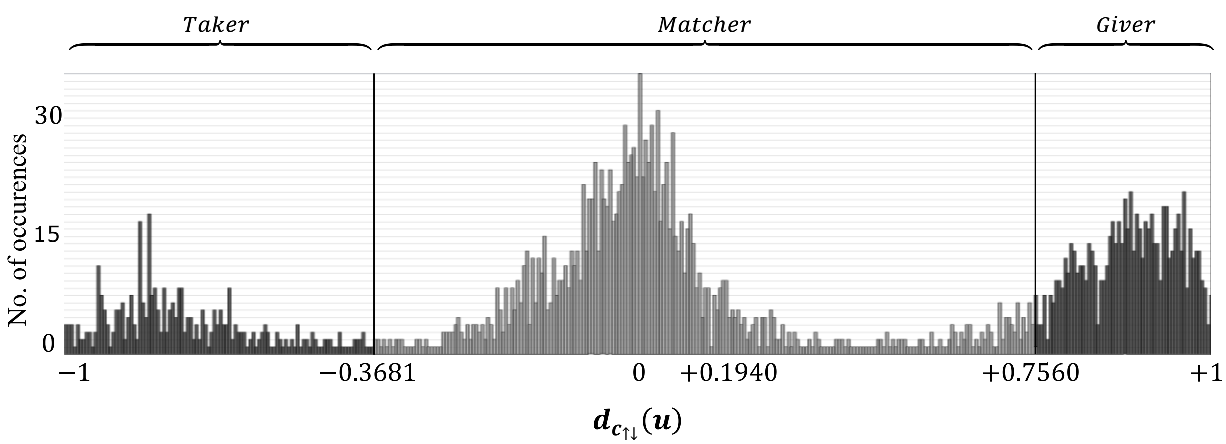

Figure 4 Distribution of Users depending on their Knowledge Contribution $\boldsymbol{d_{c_{\uparrow\downarrow}}(u)}$

With $\bar{x}_{d_{c_{\uparrow\downarrow}}(u)} = 0.1940$ and $\sigma_{d_{c_{\uparrow\downarrow}}(u)} = 0.5620$ the upper threshold of a matcher yields 0.7560 while the lower threshold results in −0.3681. The results reveal that most users in the ESN are classified as matchers (57.3%) as opposed to givers (28.5%) and takers (14.2%).

5.3 Results of the Third Step: Characterization of Users' Roles

To get deeper insights into the structural characteristics of givers, takers, and matchers we regard indegree centrality (C_i), outdegree centraliy (C_o), closeness centrality (C_c), betweenness centrality (C_b), and eigenvector centrality (C_e) for each node of the social as well as the activity graph. We furthermore investigate the sharing reciprocity (R_S) for each node of the activity graph. We find matchers to be the users with the highest centrality measures as well as the highest R_S. Table 3 shows the average values resulting for givers, takers, and matchers.

We find that, regarding the social graph, matchers are amongst the best connected users in the network showing the highest average values for all centrality measures. This means that matchers follow in average most other users (C_o) and are most often followed by others (C_i), are the closest to all other users (C_c), are most often included in the shortest paths between two other users (C_b), and have connections to other users that are themselves very well connected (C_e). As compared to this, takers are on average the least connected in the social graph concerning all centralities, closely followed by givers who manifest a slightly better connectedness than takers. Focusing on direct communication as represented by the activity graph, the results show that again, on average, the matchers are most connected while givers and takers fall behind. This holds for all centrality measures. The result for C_c indicates that these users are generally close to all other users in the activity graph and that their messages may therefore reach a large number of users in a relatively short time. At the same time, these users are most often included in the shortest paths between two other users in the activity graph of the ESN (cf. C_b) and hence are able to control or even listen to the information exchange between other users. Givers are, on average, the least connected users concerning the activity graph, with takers being only slightly more connected than givers. Regarding the exchange of sharing messages, givers and takers do not show high levels of reciprocity (cf. R_S), but matchers do. This again emphasizes the crucial role of matchers for spreading knowledge in the network and advancing the community as a whole.

	Social graph					***Activity graph***					
	C_i	C_o	C_c	C_b	C_e	C_i	C_o	C_c	C_b	C_e	R_S
Giver	0.17	0.12	3.50	0.01	2.18	0.07	0.07	0.16	0.01	0.59	25.80
Taker	0.15	0.10	3.50	0.01	2.01	0.08	0.08	0.19	0.01	0.72	5.25
Matcher	**0.45**	**0.70**	**3.52**	**0.08**	**4.10**	**0.22**	**0.21**	**0.21**	**0.13**	**4.09**	**48.90**

Table 3 Average Values for the Network Measures depending on Users' Roles (in %)

Table 4 depicts the average attributes of the different user roles. We substantiate the findings by the Social Network Analysis as we detect that a matcher writes on average 11.5 times as many messages as a giver or taker – independent of the message type (private, public, professional, unprofessional, knowledge sharing, knowledge seeking). Moreover, a matcher receives on average 7.2 times more messages as compared to a giver and 11.8 times more messages as compared to a taker respectively. Regarding the average terms per message, we detect that givers' messages contain the most (12.8 terms per message) in comparison to matchers' (10.9) and takers' (9.8). When investigating the contents of the messages exchanged more in depth, we find that givers very often share their knowledge in the form of links which inform about news in the information technology context and pass on email addresses of colleagues. They further talk about the company itself and work relevant topics (i.e. the messages broadly contain the terms "client", "projects", "management" and "team") and share presentations and internal information material as well as their experience within project work (e.g., with SAP systems). Takers predominantly search for information about the company, projects or the experience made within projects and further look for experts for project acquisition (i.e. the messages often contain the terms "look", "project", and "experience"). Matchers mainly share and demand for company and project relevant information. Yet, they also noticeably talk about the ESN itself and discuss its functionality.

In addition, when referring to other activities in the network, like the average amount of group memberships, usage of tags, bookmarks, and likes (cf. Table 4) we find that, on average, a matcher elucidates the highest participation in each aspect.

	Tags	Group memberships	Bookmarks	Likes	Sent messages	Received messages
Giver	0.05	2.12	0.03	0.26	5.82	0.24
Taker	0.02	2.07	0.02	0.25	4.64	0.39
Matcher	**0.18**	**4.80**	**0.33**	**1.60**	**60.01**	**2.78**

Table 4 Average Values for User Attributes Depending on Users' Roles

Summed up, the results of our analysis reveal that matchers take on particular importance in ESN as they are the most connected and central users concerning all centrality measures in the social graph as well as the activity graph. Moreover, they are also the most active users regarding all other activities analyzed.

Finally, the interviews with 14 users of the ESN help us to illustrate how the users reflect upon their knowledge exchanging behavior and the knowledge exchanging behavior of others. In the interviews, we found that some employees have a clear understanding of their role. For instance, one user describes himself as a taker: "I'm not an expert so I don't contribute. But I think I also do a lot of learning. I ask a lot of questions to clarify my own knowledge of certain topics. So I think my postings on Yammer are essentially to know more and its more questions than anything else." (Interview C14). Another user states that he uses Yammer to share knowledge: "[I use it] to show others how we solved problems. It's a great tool to showcase what worked and also to get feedback about what could be done better. Just the other day, I posted something and got a couple of answers – some of them pointing me in new directions, so it gave me a business benefit" (C08). This statement underlines that reciprocity enhances the motivation to participate in an ESN and therefore is a crucial aspect for its acceptance.

At the same time, a number of employees confirm that others often do not "only ask for input" (takers) or "only share knowledge" (givers) but engage in discussions where they take both sides and thus could be termed matchers. "In general it's about knowledge. Sharing knowledge and gaining knowledge. Often people use it in both ways. That's what is so great about this tool. It's easy to gain and it's easy to share." (C02). Another employee mentions the importance of matchers when it comes to appropriating the platform: "But that's the difference of users within every network. There are some more active and some less active. As long as we have enough active people, who consume but also share, the community will be sustainable" (C06). Another interviewee states: "I encourage my team as well to be on Yammer as much as possible to ensure that we don't lose that knowledge that's created out of these discussions." (C14).

Overall, the interviews most widely underpin the results of our quantitative analyses and they also show the role of Yammer as knowledge management tool.

6 Discussion

While prior studies mainly focus on structural characteristics (e.g., number of likes) when identifying users' roles based on their knowledge contribution in ESN, they do not consider the content perspective in sufficient detail [24,25,72,73]. Against this background, we propose and apply an approach consisting of three steps that allows us to distinguish between givers, takers, and matchers based on their knowledge contribution via ESN messages. Our results illustrate that the contents necessarily need to be considered in order to get reliable results as the mere consideration of the structural characteristics may lead to misinterpretations of the results (cf. also 6.1).

6.1 Matchers as Central Element of an ESN

Our results show that the clear majority of the users in the ESN act as matchers, i.e. they are willing to help others but as part of a "this for that"-approach also want to get something back. This is underscored by interviews in which users of the considered ESN shared their observations that people use Yammer to gain and to share, as it is easy to do both. Beyond that, we find that matchers play a central role in ESN as they keep the network alive due to their high network interconnectedness and activity. They are by far the best connected in the social as well as the activity graph and are also the most active concerning all other activities analyzed (such as giving likes, bookmarks etc.). These results elucidate that matchers are the most important users in the ESN as they connect the users and spread the information in the network. The fact that they receive most messages means that by demanding information from users (in turn for sharing information), they encourage other users to also participate in the ESN, and thus they keep the network alive and together. They can for instance contribute to bridging structural holes [112] between sub-networks in the ESN which do not or only little overlap (due to their high C_b) and moreover, they may enable a more effective and rapid information exchange between different working groups that are for instance only sparsely connected. More generally speaking, matchers are crucial for the diffusion of innovative ideas which essentially depends on how people are connected and influence each other [113].

Referring to reciprocity, matchers are also those users who mostly give knowledge back to the community in return for help, which illustrates that they aim at bringing forward the community as a whole. This is in line with Wasko and Faraj [52] who found that giving back to a community in return for help is by far the most cited reason why people participate. It also complies with Kollock [114], who states that people helping others indeed hope to receive something back in return. He argues that these people expect interaction to be available in the future and therefore, the possibility of future reciprocation must be given. It can be concluded that companies should use technologies that show the identities of the users and archive discussions in a searchable format.

In addition to sending most messages, matchers also write in average the longest messages which indicates that they put a lot of effort in the ESN. For this part, our findings are in line with Berger et al. [74] who show that users adding value for others are amongst the best connected users in ESN. In addition, referring to the contents exchanged, matchers discuss more about the network itself as well as its functionalities, as compared to givers and takers who rather exchange work related information. This fosters that matchers are the central element of the ESN keeping the network alive and developing it further. Moreover, our results reveal that 50% of the messages are written by approximately 1% of the users, whereby all of these are matchers. This complies with Nielsen [115] who find that only 10% of all users in a social community create 100% of its content as well as Trier and Richter [72] who state that a smaller group of information contributors in organizational networks competes for a large group of retrievers in order to grow their topic. Moreover, also Yardi et al. [20] come to know that only a few individuals receive the majority of the attention in ESN. In contrast to other studies, our results elucidate that an investigation of the contents is crucial to reach valid results. Contrary to Cetto et al. [24] who base their identification of users' roles on the relation of read and write accesses (without considering the contents), we illustrate that solely regarding the number of messages is not sufficient. When identifying matchers based on their contents exchanged, we find that they also write most of the messages. Regarding merely the number of messages exchanged, these users would subsequently be identified as givers. Hence, understanding the contents exchanged more in depth is crucial for a reasonable identification of users' roles.

6.2 Givers and Takers as Less Participating Users

Surprisingly, givers and takers are by comparison less participating users. They write and receive rather few messages and are not as well connected in the social and activity graph as compared to matchers. Both, givers and takers could have been anticipated to be more active and better connected. Givers could have been expected to be more active through a higher absolute amount of outgoing messages and a better connectivity while takers could have been expected to gain more knowledge from the network through a higher absolute amount of incoming messages and a better connectivity respectively. A reason for the rather low connectivity of givers concerning the activity graph can be that the activity graph concerns private communication while givers might prefer to share their knowledge not in a private but rather in a public context with the aim to reach as many users as possible. But as opposed to this, givers also have comparably few group memberships. This is rather surprising as groups could be used to reach multiple users with only a single message and thus offer a good opportunity to spread knowledge more easily. Consequently, as matchers communicate most in both – the private as well as the public context – companies are well advised to precisely identify and address their matchers in order to support an effective and successful exchange of knowledge within the organization.

Nevertheless, it cannot be neglected that also givers and takers are of certain importance for the company in such that givers also have the potential to spread knowledge by giving their knowledge to others while takers are important in the sense that they can gain new work relevant knowledge through asking questions and participating in the ESN. This is in line with Beck et al. [25] who provide evidence that the mix of questions and answers in communications impacts the quality of knowledge exchanged. Hence, also these users should be encouraged to take an active part in the ESN, for instance through incentives such as a bonus for a certain participation rate in the ESN.

Our results furthermore reveal that a high amount of users is enrolled but has not even one written nor received message in the ESN at all. These so called "lurkers" make up to 63.2% of all enrolled users in our dataset. This is in keeping with Katz [116] who states that the majority (up to even 90%) of online community members can be identified as lurkers. Nonnecke and Preece [68] analyze the reasons for lurking in online communities and reveal that amongst the main reasons for this behavior are privacy and safety concerns, reluctance, and usability problems [69]. This is in line with the results delivered in our word clouds which show that users are still unsure how to properly use the platform and are not completely convinced of it. This leads to the assumption that these aspects can really be a problem for some users which prevent them from participating in the ESN and which in turn results in lurking.

6.3 Characteristics Across All Users' Roles

Across all users' roles, we find that employees use the ESN primarily for professional purposes. Moreover, the majority of the professional messages are knowledge sharing which shows that the users are generally cooperative and willing to share their professional knowledge with and thus help other users. Hence, they use the network as communication channel which enables them to spread their knowledge more easily with a vast amount of people (as compared to offline communication). This is in line with Kane [5] and Aral et al. [3] who argue that social media in the organizational context support and fundamentally change the way people communicate, collaborate, consume, and create.

Moreover, the fact that the majority of messages are intended to share knowledge can lead to the assumption that only one knowledge sharing message as answer to a knowledge seeking message may not be sufficient for the explanation of certain circumstances. Users often need more than one message to explain or discuss certain aspects in depth, which in turn leads to a higher amount of knowledge sharing messages. This enriches the network as discussions can generate new knowledge and encourage other users to give their opinion and thus also share their knowledge within the network.

Moreover, regarding the message contents, our results reveal that knowledge sharing messages tend to point to helpful links and tags which in turn ensures that users find information more easily. Further, apart from work related information, knowledge sharing messages comprise information about the network (Yammer) making clear that the functionalities of the network itself are in focus of communication and need further clarification.

Our results also show that knowledge seeking messages often thank users for messages, which can indicate that the message was an answer to a knowledge sharing message following a previous question. A reason for the prevalence of sharing messages can be that the employees of the consulting company see the ESN as a chance to stand out from the crowd and promote themselves. Through answering questions and sharing links they can show that they own a lot of knowledge and are experts in their fields. Against this backdrop, they increase their visibility in the company and might be recruited for more projects which in turn enhances their reputation and can speed-up their career path in the company. This illustrates that ESN enable companies to better detect and trace their experts which then again leads to a more efficient project staffing. This is in accordance with Berger et al. [74], who state that companies are well advised to identify their key users in ESN to enable a more effective and rapid exchange of information between different working groups. Therefore, ESN providers should better invest in the provision of analytic functions to improve the visibility of the most important users in the network.

6.4 Limitations and Future Research

Even though our research provides first interesting insights into the identification of as well as the characteristics of givers, takers, and matchers in ESN, there are several limitations which can serve as starting points for future research.

First, we only considered one single company, which provided us with the relevant user and message data needed to conduct this research. Nevertheless, the ESN was actively used by a large number of users for sharing and seeking knowledge and we assume that our findings also hold for other companies using ESN.

Second, for message classification, we defined knowledge sharing messages as those messages which contain helpful information or offer help to other users and we defined knowledge seeking messages as those messages containing signs that the user receives information, demands for information, or demands help from other users. Obviously, this definition cannot hold in all cases as messages can contain parts from both definitions. However, the classification algorithm used provides probabilities that a message pertains to the knowledge sharing or knowledge seeking class and thus provides the information that a message contains comparably more knowledge seeking or more knowledge sharing content. Hence, it may be well assumed these probabilities are appropriate for being used in our context.

Third, we classified users as givers, takers, and matchers based on the content of their messages. In so doing, we focused on relevant words in messages and accepted a loss of semantics as we selected and filtered words, ignored verbosity (part of speech), and the context of a message in the message flow. However, we assume that the remaining relevant words represent the main features for knowledge contribution in ESN. Further in-depth analysis regarding the context of the related thread or group in which a message is written is needed to include the message flows in the proposed approach and answer the question whether a knowledge seeking message is often followed by a knowledge sharing message. Furthermore, it would be of interest to analyze chronological orders of messages, for instance whether a discussion is started and finished by a knowledge seeking message while in between the discussion is dominated by knowledge sharing messages.

Moreover, we did not consider the time factor of the messages and the users' roles. While in a first step, it seemed appropriate to take such a static perspective, further studies are needed to analyze this aspect in-depth. It would be promising to analyze the time-based change of users' roles (e.g., from takers to matchers to givers) and the users' life cycles. In the course of this development it would also be interesting to incorporate further characteristics of the users of each role beyond the social embeddedness (e.g. demographics, position in the organization and hierarchies) in order to get a comprehensive picture of givers, takers, and matchers.

7 Conclusion

Despite emerging scientific work in the field of ESN, we still observe a lack of research on employees' knowledge exchange practices in ESN, for instance how users contribute knowledge in ESN. While there is a growing body of literature on identifying users' roles which mainly considers the users' structural characteristics [24,25,72–74], the content perspective for getting further insights in the knowledge contribution of users as well as their user roles respectively, is still widely unexplored.

Thus, the aim of this paper is to investigate how users' roles in ESN can be identified based on users' knowledge contribution in messages. Further, we determine how users can be characterized based on their roles and the messages that they exchanged with others. We propose an approach consisting of three steps that allows us to distinguish between givers, takers, and matchers based on their knowledge contribution via ESN messages.

The application of our approach to a dataset of a large multinational consulting company delivers exciting results which elucidate the importance for companies to also consider ESN as sources of company relevant knowledge. First we find that users wrote mainly professional messages and aimed at sharing their knowledge which shows that the users are generally cooperative and willing to help others. In addition to that, our results reveal that most users in the ESN can be identified as matchers and therefore, they offer and deliver information to others, but also demand information in return. Moreover, they are amongst the best connected, most active users in the network, giving them a central position in the ESN as they keep the network together and alive. Therefore, organizations are well recommended to identify and address their matchers for an effective knowledge management within the organization.

From a theoretical perspective, our findings contribute to the development of a more refined understanding of ESN usage in knowledge-intensive work. Through identifying users' roles we shed light on their networking behavior and help to better understand their characteristics. As such our study builds upon research on users' roles in knowledge contribution in ESN and extends its line of thoughts by using a content perspective. From a practical point of view, our insights can help companies to better understand the knowledge contribution behavior of their employees in ESN. Companies are well advised to better investigate and to become acquainted with the characteristics of their ESN users to ensure an efficient knowledge management in the company. Our study can support companies to attain this aim as it offers valuable insights in the knowledge contribution behavior of their ESN users.

Summing up, we believe that our study is a first but indispensable step in terms of studying users' knowledge contribution in ESN messages. We hope that our results will stimulate further research on that fascinating topic and support practitioners to better understand and use ESN for knowledge management.

References

[1] Thompson, V., "Worldwide Enterprise Social Networks and Online Communities 2015 – 2019 Forecast and 2014 Vendor Shares," https://www.re-imaginework.com/sites/default/files/public/document/IDC%20-%20Worldwide%20Enterprise%20Social%20Networks%20and%20Online%20Communities%202015%20-%202019%20Forecast%20and%202014%20Vendor%20Shares.pdf.

[2] Aoun, C. and Vatanasakdakul, S., "Social Media in the Workplace: Key Drivers for Inclusive Innovation," in *Proceedings of the Eighteenth Americas Conference on Information Systems (AMCIS)*, Seattle, WA, 2012.

[3] Aral, S., Dellarocas, C., and Godes, D., "Introduction to the Special Issue - Social Media and Business Transformation: A Framework for Research," *Information Systems Research*, vol. 24, no. 1, pp. 3–13, 2013.

[4] von Krogh, G., "How Does Social Software Change Knowledge Management? Toward a Strategic Research Agenda," *The Journal of Strategic Information Systems*, vol. 21, no. 2, pp. 154–164, 2012.

[5] Kane, G.C., "The Evolutionary Implications of Social Media for Organizational Knowledge Management," *Information and Organization*, vol. 27, no. 1, pp. 37–46, 2017.

[6] Steinhueser, M., Richter, A., and Smolnik, S., "How to Bridge the Boundary? Determinants of Inter-Organizational Social Software Usage," *Electronic Markets*, vol. 25, no. 4, pp. 267–281, 2015.

[7] Meske, C. and Stieglitz, S., "Adoption and Use of Social Media in Small and Medium-Sized Enterprises," in *Working Conference on Practice-Driven Research on Enterprise Transformation*, Springer, Heidelberg, 2013.

[8] Richter, D., Riemer, K., and vom Brocke, J., "Internet Social Networking – Research State of the Art and Implications for Enterprise 2.0," *Business & Information Systems Engineering*, vol. 3, no. 2, pp. 89–101, 2011.

[9] Riemer, K., Stieglitz, S., and Meske, C., "From Top to Bottom: Investigating the Changing Role of Hierarchy in Enterprise Social Networks," *Business & Information Systems Engineering*, vol. 57, no. 3, pp. 197–212, 2015.

[10] Turban, E., Bolloju, N., and Liang, T.-P., "Enterprise Social Networking: Opportunities, Adoption, and Risk Mitigation," *Journal of Organizational Computing and Electronic Commerce*, vol. 21, no. 3, pp. 202–220, 2011.

[11] Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., and Venkatraman, V., "Visions and Voices on Emerging Challenges in Digital Business Strategy," *MIS quarterly*, vol. 37, no. 2, pp. 633–661, 2013.

[12] Richter, A. and Riemer, K., "Malleable End-User Software," *Business & Information Systems Engineering*, vol. 5, no. 3, pp. 195–197, 2013.

[13] Herzog, C., Richter, A., Steinhüser, M., Hoppe, U., and Koch, M., "Methods And Metrics For Measuring The Success Of Enterprise Social Software-What We Can Learn From Practice And Vice Versa," in *Proceedings of the Twenty-First European Conference on Information Systems (ECIS)*, Utrecht, Netherlands, 2013.

[14] Phang, C.W., Kankanhalli, A., and Sabherwal, R., "Usability and Sociability in Online Communities: A Comparative Study of Knowledge Seeking and Contribution," *Journal of the American Society for Information Science and Technology*, vol. 10, no. 10, pp. 721–747, 2009.

[15] Schroer, J. and Hertel, G., "Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians And Why They Do It," *Media Psychology*, vol. 12, no. 1, pp. 96–120, 2009.

[16] Kankanhalli, A., Tan, B., and Wei, K.-K., "Understanding seeking from electronic knowledge repositories: An empirical study," *Journal of the American Society for Information Science and Technology*, vol. 56, no. 11, pp. 1156–1166, 2005.

[17] Wasko, M.M. and Faraj, S., "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS quarterly*, vol. 29, no. 1, pp. 35–57, 2005.

[18] Zhang, X. and Wang, C., "Network Positions and Contributions to Online Public Goods: The Case of Chinese Wikipedia," *Journal of Management Information Systems*, vol. 29, no. 2, pp. 11–40, 2012.

[19] Yan, Y. and Davison, R.M., "Exploring Behavioral Transfer from Knowledge Seeking to Knowledge Contributing: the Mediating Role of Intrinsic Motivation," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 6, pp. 1144–1157, 2013.

[20] Yardi, S., Golder, S.A., and Brzozowski, M.J., "Blogging at Work and the Corporate Attention Economy," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009.

[21] Muller, M., Shami, N.S., Millen, D.R., and Feinberg, J., "We Are All Lurkers: Consuming Behaviors among Authors and Readers in an Enterprise File-sharing Service," in *Proceedings of the Sixteenth ACM International Conference on Supporting Group Work*, 2010.

[22] Yeow, A., Johnson, S., and Faraj, S., "Lurking: Legitimate or Illegitimate Peripheral Participation?," in *Proceedings of the Twenty-seventh International Conference on Information Systems (ICIS)*, Milwaukee, WI, 2006.

[23] Tan, A.-H., "Text Mining: Promises and Challenges," *South East Asia Regional Computer Confederation, Singapore*, 1999.

[24] Cetto, A., Klier, J., Klier, M., Richter, A., and Wiesneth, K., "The Blessing of Giving: Knowledge Sharing and Knowledge Seeking in Enterprise Social Networks," in *Proceedings of the Twenty-Fourth European Conference on Information Systems (ECIS)*, Istanbul, Turkey, 2016.

[25] Beck, R., Pahlke, I., and Seebach, C., "Knowledge Exchange and Symbolic Action in Social Media-Enabled Electronic Networks of Practice: A Multilevel Perspective on Knowledge Seekers and Contributors," *MIS quarterly*, vol. 38, no. 4, pp. 1245–1270, 2014.

[26] Denyer, D., Parry, E., and Flowers, P., ""Social", "open" and "participative"? Exploring personal and organizational effects of Enterprise 2.0 use," *Long Range Planning*, vol. 44, 5-6, pp. 375–396, 2011.

[27] Holtzblatt, L.J., Damianos, L.E., and Weiss, D., "Factors Impeding Wiki Use in the Enterprise," in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 2010.

[28] Ip, R.K.F. and Wagner, C., "Weblogging: A Study of Social Computing and Its Impact on Organizations," *Decision Support Systems*, vol. 45, no. 2, pp. 242–250, 2008.

[29] Overfeld, P., Scifleet, P., Richter, A., and Riemer, K., "Eliciting the Anatomy of Technology Appropriation Processes: A Case Study in Enterprise Social Media," in *Proceedings of the Twentieth European Conference on Information Systems (ECIS)*, Barcelona, Spain, 2012.

[30] DiMicco, J.M., Geyer, W., Millen, D.R., Dugan, C., and Brownholtz, B., "People Sensemaking and Relationship Building on an Enterprise Social Network Site," in *Proceedings of the Forty-Second Annual Hawaii International Conference on System Sciences*, Waikoloa, Big Island, HI, 2009.

[31] Brzozowski, M.J., "WaterCooler: : Exploring an Organization through Enterprise Social Media," in *Proceedings of the 2009 International ACM Conference on Supporting Group Work*, pp. 219–228, Sanibel Island, FL, 2009.

[32] Thom-Santelli, J., Millen, D.R., and Gergle, D., "Organizational Acculturation and Social Networking," in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, ACM, 2011.

[33] Wu, L., "Social Network Effects on Productivity and Job Security: Evidence from the Adoption of a Social Networking Tool," *Information Systems Research*, vol. 24, no. 1, pp. 30–51, 2013.

[34] Bobsin, D. and Hoppen, N., "The Structure of Organizational Virtual Social Networks," in *Proceedings of the Sixth International Conference on Information Resources Management*, 2013.

[35] Behrendt, S., Klier, J., Klier, M., Richter, A., and Wiesnet, K., "The Impact of Formal Hierarchies on Enterprise Social Networking Behavior," in *Proceedings of the Thirty-Sixth International Conference on Information Systems (ICIS)* , Fort Worth, USA, 2015.

[36] Koo, C., Wati, Y., and Jung, J.J., "Examination of How Social Aspects Moderate the Relationship between Task Characteristics and Usage of Social Communication Technologies (SCTs) in Organizations," *International Journal of Information Management*, vol. 31, no. 5, pp. 445–459, 2011.

[37] Kuegler, M. and Smolnik, S., "Uncovering the Phenomenon of Employees' Enterprise Social Software Use in the Post-Acceptance Stage - Proposing a Use Typology," in *Proceedings of the Twenty-Second European Conference on Information Systems (ECIS)*, 2014.

[38] Stieglitz, S., Riemer, K., and Meske, C., "Hierarchy or Activity? The Role of Formal and Informal Influence in Eliciting Responses from Enterprise Social Networks," in *Proceedings of the Twenty-Second European Conference on Information Systems (ECIS)*, 2014.

[39] Ortbach, K. and Recker, J., "Do Good Things and Talk About Them: A Theory of Academics Usage of Enterprise Social Networks for Impression Management Tactics," in *Proceedings of the Thirty-Fifths International Conference on Information Systems*, Auckland, 2014.

[40] Recker, J. and Lekse, D., "Spatial Preferences in Microblogging," *Journal of Information Technology,* Forthcoming, 2015.

[41] Grant, A., *Give and Take: a Revolutionary Approach to Success*, Hachette UK, 2013.

[42] Bock, G.-W., Zmud, R.W., Kim, Y.-G., and Lee, J.-N., "Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Forces, and Organizational Climate," *MIS quarterly*, pp. 87–111, 2005.

[43] Kankanhalli, A., Tan, B., and Wei, K.-K., "Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation," *MIS quarterly*, vol. 29, no. 1, pp. 113–143, 2005.

[44] Krogh, G. von, "Care in Knowledge Creation," *California management review*, vol. 40, no. 3, pp. 133–153, 1998.

[45] Newman, M., *Networks: An Introduction*, Oxford University Press, 2010.

[46] Cho, H., Chen, M., and Chung, S., "Testing an Integrative Theoretical Model of Knowledge-Sharing Behavior in the Context of Wikipedia," *Journal of the Association for Information Science and Technology*, vol. 61, no. 6, pp. 1198–1212, 2010.

[47] Sun, Y., Fang, Y., and Lim, K.H., "Understanding Sustained Participation in Transactional Virtual Communities," *Decision Support Systems*, vol. 53, no. 1, pp. 12–22, 2012.

[48] Jackson, A., Yates, J., and Orlikowski, W., "Corporate Blogging: Building Community Through Persistent Digital Talk," in *Proceedings of the Fortieth Hawaii International Conference on System Sciences (HICSS)*, 2007.

[49] Hasan, H. and Pfaff, C.C., "The Wiki: An Environment to Revolutionise Employees' Interaction With Corporate Knowledge," in *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, ACM, 2006.

[50] Kim, S.T., Lee, C.K., and Hwang, T., "Investigating the Influence of Employee Blogging on IT Workers' Organisational Citizenship Behaviour," *International Journal of Information Technology and Management*, vol. 7, no. 2, pp. 178–189, 2008.

[51] Thom-Santelli, J., Muller, M.J., and Millen, D.R., "Social Tagging Roles: Publishers, Evangelists, Leaders," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2008.

[52] Wasko, M.M. and Faraj, S., ""It Is What One Does": Why People Participate and Help Others in Electronic Communities of Practice," *The Journal of Strategic Information Systems*, vol. 9, no. 2, pp. 155–173, 2000.

[53] Nowak, M.A. and Sigmund, K., "Evolution of Indirect Reciprocity," *Nature*, vol. 437, no. 7063, pp. 1291–1298, 2005.

[54] Mathews, M.A. and Green, J.D., "Looking at Me, Appreciating You: Self-Focused Attention Distinguishes Between Gratitude and Indebtedness," *Cognition & Emotion*, vol. 24, no. 4, pp. 710–718, 2010.

[55] Chai, S., Das, S., and Rao, H.R., "Factors Affecting Bloggers' Knowledge Sharing: An Investigation Across Gender," *Journal of Management Information Systems*, vol. 28, no. 3, pp. 309–342, 2011.

[56] Riemer, K. and Richter, A., "Tweet Inside: Microblogging in a Corporate Context," 2010.

[57] Cleveland, S., "Using Microblogging for Lessons Learned in Information Systems Projects," in *Proceedings of the Seventh International Research Workshop on Information Technology Project*, Orlando, FL, 2012.

[58] Cleveland, S. and Ellis, T., "The Role of Microblogging Capacities in Knowledge Sharing and Collaboration in Virtual Teams," in *IEEE Proceedings of the Twenty-First Americas Conference on Information Systems*, Puerto Rico, USA, 2015.
[59] Zhang, J., Qu, Y., Cody, J., and Wu, Y., "A Case Study of Microblogging in the Enterprise: Use, Value, and Related Issues," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2010.
[60] Ebner, M., Lienhardt, C., Rohs, M., and Meyer, I., "Microblogs in Higher Education – A Chance to Facilitate Informal and Process-Oriented Learning?," *Computers & Education*, vol. 55, no. 1, pp. 92–100, 2010.
[61] Zhao, D., Rosson, M.B., Matthews, T., and Moran, T., "Microblogging's Impact on Collaboration Awareness: A Field Study of Microblogging Within and Between Project Teams," in *International Conference on Collaboration Technologies and Systems (CTS)*, IEEE, 2011.
[62] Trusov, M., Bodapati, A.V., and Bucklin, R.E., "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, vol. 47, no. 4, pp. 643–658, 2010.
[63] Drucker, P.F., *Landmarks of Tomorrow: A Report on the New 'Post-Modern' World*, Harper & Row, New York, 1959.
[64] Emanoil, M., Alexandra, S., and Mihaela, C.L., "Knowledge Workers – Drivers to Organizational Performance in a Knowledge-Based Economy," *Mathematical Methods in Engineering and Economics*, pp. 150–154, 2014.
[65] Parise, S., Cross, R., and Davenport, T.H., "Strategies for Preventing a Knowledge-Loss Crisis," *MIT Sloan Management Review*, vol. 47, no. 4, p. 31, 2006.
[66] Reinhardt, W., Schmidt, B., Sloep, P., and Drachsler, H., "Knowledge Worker Roles and Actions — Results of Two Empirical Studies," *Knowledge and Process Management*, vol. 18, no. 3, pp. 150–174, 2011.
[67] Yuqing, R., Kraut, R., and Kiesler, S., "Applying Common Identity and Bond Theory to Design of Online Communities," *Organization Studies*, vol. 28, no. 3, pp. 377–408, 2007.
[68] Nonnecke, B. and Preece, J., "Lurker Demographics: Counting the Silent," in *Proceedings of the SIGCHI Conference on Human Factos in Computing Systems*, The Hague, Netherlands, 2000.
[69] Nonnecke, B. and Preece, J., "Why Lurkers Lurk," in *Proceedings of the Seventh Americas Conference on Information Systems (AMCIS)*, Boston, MA, 2001.
[70] Schneider, A., Krogh, G. von, and JäGer, P., ""What's Coming Next?" Epistemic Curiosity and Lurking Behavior in Online Communities," *Computers in Human Behavior*, vol. 29, no. 1, pp. 293–303, 2013.
[71] Gloor, P.A., Laubacher, R., Dynes, S.B.C., and Zhao, Y., "Visualization of Communication Patterns in Collaborative Innovation Networks-Analysis of Some W3C Working Groups," in *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pp. 56–60, ACM, 2003.
[72] Trier, M. and Richter, A., "The Deep Structure of Organizational Online Networking - An Actor-Oriented Case Study," *Information Systems Journal*, vol. 25, no. 5, pp. 465–488, 2015.
[73] Holtzblatt, L., Drury, J.L., Weiss, D., Damianos, L.E., and Cuomo, D., "Evaluating the Uses and Benefits of an Enterprise Social Media Platform," *Journal of Social Media for Organizations*, vol. 1, no. 1, pp. 1–21, 2013.
[74] Berger, K., Klier, J., Klier, M., and Richter, A., ""Who is Key…?"- Value Adding Users in Enterprise Social Networks," in *Proceedings of the Twenty-Second European Conference on Information Systems (ECIS)*, 2014.
[75] Richter, A., Heidemann, J., Klier, M., and Behrendt, S., "Success Measurement of Enterprise Social Networks," in *Wirtschaftsinformatik Proceedings 2013*, Leipzig, Germany, 2013.
[76] Park, S. and An, D.U., "Automatic E-Mail Classification Using Dynamic Category Hierarchy and Semantic Features," *IETE Technical Review*, vol. 27, no. 6, pp. 478–492, 2010.
[77] Abrahams, A.S., Jiao, J., Wang, G.A., and Fan, W., "Vehicle Defect Discovery From Social Media," *Decision Support Systems*, vol. 54, no. 1, pp. 87–97, 2012.

[78] Li, N. and Wu, D.D., "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast," *Decision Support Systems*, vol. 48, no. 2, pp. 354–368, 2010.

[79] Dumais, S., Platt, J., Heckerman, D., and Sahami, M., "Inductive Learning Algorithms and Representations for Text Categorization," in *Proceedings of the Seventh International Conference on Information and Knowledge Management*, ACM, 1998.

[80] Huang, X. and Croft, W.B., "A Unified Relevance Model for Opinion Retrieval," in *Proceedings of the Eighteenth ACM Conference on Information and Knowledge Management*, ACM, 2009.

[81] Miles, M.B. and Huberman, A.M., *Qualitative Data Analysis: An Expanded Sourcebook*, SAGE Publications, Thousand Oaks, CA, 1994.

[82] Krippendorff, K., "Estimating the Reliability, Systematic Error and Random Error of Interval Data," *Educational and Psychological Measurement*, vol. 30, no. 1, pp. 61–70, 1970.

[83] Hayes, A.F. and Krippendorff, K., "Answering the Call for a Standard Reliability Measure for Coding Data," *Communication Methods and Measures*, vol. 1, no. 1, pp. 77–89, 2007.

[84] Storn, R. and Price, K., "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization Over Continuous Spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[85] Tralli, V., Vaccari, A., Verdone, R., and Andrisano, O., "Adaptive Time and Frequency Resource Assignment with COFDM for LMDS Systems," *IEEE Transactions on Communications*, vol. 49, no. 2, pp. 235–238, 2001.

[86] Porter, M.F., "An Algorithm for Suffix Stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[87] Sebastiani, F., "Machine Learning in Automated Text Categorization," *ACM Computing Surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002.

[88] Wallach, H.M., "Topic Modeling: Beyond Bag-of-Words," in *Proceedings of the Twenty-third International Conference on Machine Learning*, ACM, 2006.

[89] Hua, J., Tembe, W.D., and Dougherty, E.R., "Performance of Feature-Selection Methods in the Classification of High-Dimension Data," *Pattern Recognition*, vol. 42, no. 3, pp. 409–424, 2009.

[90] Quinlan, J.R., "Induction of Decision Trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[91] Cortes, C. and Vapnik, V., "Support-Vector Networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[92] Cover, T. and Hart, P., "Nearest Neighbor Pattern Classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[93] John, G.H. and Langley, P., "Estimating Continuous Distributions in Bayesian Classifiers," in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338–345, Morgan Kaufmann Publishers Inc., 1995.

[94] Park, D.C., El-Sharkawi, M.A., Marks, R.J., Le Atlas, and Damborg, M.J., "Electric Load Forecasting Using an Artificial Neural Network," *IEEE transactions on Power Systems*, vol. 6, no. 2, pp. 442–449, 1991.

[95] Forman, G., "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *Journal of machine learning research*, vol. 3, Mar, pp. 1289–1305, 2003.

[96] Sokolova, M., Japkowicz, N., and Szpakowicz, S., "Beyond Accuracy, F-score and ROC: A Family of Discriminant Measures for Performance Evaluation," in *Australasian Joint Conference on Artificial Intelligence*, pp. 1015–1021, Springer Berlin, Heidelberg, 2006.

[97] Kohavi, R., "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *IJCAI*, vol. 14, no. 2, pp. 1137–1145, 1995.

[98] Platt, J., "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," in *Advances in Large Margin Classifiers,* A. J. Smola, Ed., vol. 10, pp. 61–74, MIT press, 2000.

[99] Wasserman, S. and Faust, K., *Social Network Analysis*: *Methods and Applications*, Cambridge Univ. Press, Cambridge, UK, 2009.

[100] Krasnova, H., Koroleva, K., and Veltri, N.F., "Investigation of the Network Construction Behavior on Social Networking Sites," in *Proceedings of the Thirty-First International Conference on Information Systems (ICIS)*, St. Louis, MO, 2010.

[101] Freeman, L.C., "Social Network Analysis: Definition and History," in *Encyclopedia of Psychology,* A. E. Kazdin, Ed., pp. 350–351, Oxford University Press, London, 2000.

[102] Freeman, L.C., "Centrality in Social Networks Conceptual Clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[103] Bonacich, P., "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.

[104] Adamic, L.A. and Adar, E., "Friends and Neighbors on the Web," *Social networks*, vol. 25, no. 3, pp. 211–230, 2003.

[105] Bampo, M., Ewing, M.T., Mather, D.R., Stewart, D., and Wallace, M., "The Effects of the Social Structure of Digital Networks on Viral Marketing Performance," *Information Systems Research*, vol. 19, no. 3, pp. 273–290, 2008.

[106] Heidemann, J., Klier, M., and Probst, F., "Identifying Key Users in Online Social Networks: A PageRank Based Approach," in *Proceedings of the Thirty-First International Conference on Information Systems (ICIS)*, St. Louis, MO, 2010.

[107] Gouldner, A.W., "The Norm of Reciprocity: A Preliminary Statement," *American sociological review*, pp. 161–178, 1960.

[108] Snijders, T.A.B., van de Bunt, G.G., and Steglich, C.E.G., "Introduction to Stochastic Actor-Based Models for Network Dynamics," *Social networks*, vol. 32, no. 1, pp. 44–60, 2010.

[109] Contractor, N.S., Wasserman, S., and Faust, K., "Testing Multitheoretical, Multilevel Hypotheses About Organizational Networks: An Analytic Framework and Empirical Example," *Academy of Management Review*, vol. 31, no. 3, pp. 681–703, 2006.

[110] Behrendt, S., Richter, A., and Trier, M., "Mixed Methods Analysis of Enterprise Social Networks," *Computer Networks*, vol. 75, pp. 560–577, 2014.

[111] Richter, A. and Riemer, K., "The Contextual Nature of Enterprise Social Networking: A Multi Case Study Comparison," in *Proceedings of the Twenty-First European Conference on Information Systems (ECIS)*, Utrecht, Netherlands, 2013.

[112] Burt, R.S., *Structural Holes: The Social Structure of Competition*, Harvard University Press, Cambridge, MA, 1992.

[113] Ciriello, R., Hu, D., and Schwabe, G., "Identifying Patterns of Idea Diffusion in Innovator Networks," in *Proceedings of the Thirty-Fourth International Conference on Information Systems (ICIS)*, Milan, Italy, 2013.

[114] Kollock, P., *The Economies of Online Cooperation*: *Gifts and Public Goods in Cyberspace*, London: Routledge, Smith, M.A., 1999.

[115] Nielsen, J., "The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities," https://www.nngroup.com/articles/participation-inequality/.

[116] Katz, J., "Luring the Lurkers," https://news.slashdot.org/story/98/12/28/ 1745252/luring-the-lurkers.